BAB II

TINJAUAN PUSTAKA

II.1 Penelitian Terkait

Dalam memudahkan proses penelitian, peneliti perlu melakukan kajian terhadap penelitian sebelumnya yang memiliki keterkaitan dengan judul penelitian. Kajian pustaka ini bertujuan untuk menghindari pengulangan kata-kata atau kalimat yang sama dengan penelitian sebelumnya. Berikut adalah hasil dari penelitian-penelitian terdahulu yang telah disertakan sebagai berikut:

Pada penelitian yang dilakukan oleh(Ashok et al., 2023) dengan judul penelitian "Building a Medium Scale Dataset for Nondestructive Disease Classification in Mango Fruits Using Machine Learning and Deep Learning Models". Permasalahan pada penelitian ini kualitas buah-buahan sering dinilai berdasarkan atribut sensorik seperti rasa dan aroma, bentuk, ukuran, warna. Dengan menggunakan pengolahan citra komputer yang dikombinasikan dengan algoritma pembelajaran. Penelitian ini bertujuan untuk menguji keandalan dataset yang diusulkan dengan menggunakan pendekatan baru berdasarkan analisis fungsi diskriminan (discriminant function analysis, DFA) peneliti melakukan analisis benchmark yang ekstensif menggunakan algoritma statistik dan pembelajaran dalam seperti support vector machine (SVM) dan convolutional neural network (CNN) pada dataset yang telah divalidasi. Hasil dari penelitian menunjukkan bahwa SVM mencapai akurasi klasifikasi buah mangga penyakit yang signifikan sebesar 95%, sementara CNN sebesar 91,52%.

Peneliti mengusulkan dataset dapat digunakan oleh penelitian lain untuk penelitian lebih lanjut dalam evaluasi kualitas buah mangga menggunakan teknik pemrosesan citra dan pembelajaran mesin.

Penelitian oleh(X. Yang et al., 2019) dengan judul penelitian "Machine learning for cultivar classification of apricots (Prunus armeniaca L.) based on shape features". Penelitian ini bertujuan untuk klasifikasi varietas aprikot menggunakan machine learning berdasarkan fitur bentuk (shape features). Identifikasi varietas apriko menggunakan enam metode machine learning yaitu decision tree, K-nearest neighbour, naive Bayes, linear discriminant analysis, support vector machine, dan back propagation neural network. Berdasarkan hasil penelitian menunjukkan varietas aprikot memiliki perbedaan signifikan dalam fitur bentuk dengan klasifikasi support vector machine algoritma memberikan hasil terbaik tingkat akurasi 90.7% dalam dataset pengujian.

Penelitian selanjutnya oleh(Maghfirah & Mohamad, 2022) dengan judul penelitian "Klasisikasi Kematangan tomat dengan model warna yang berbeda menggunakan linear discriminant analisis (LDA)". Penelitian ini bertujuan untuk mengklasifikasikan kematangan tomat menggunakan gambar dominan berbasis warna. Dalam penelitian tersebut menggunakan tiga model warna *HSV, YcbCr* dan *CIElab* untuk menguji kematangan buah tomat berdasarkan warna, data yang digunakan gambar berjumlah 126 kemudian dibagi menjadi dua kelas matang dan mentah. Berdasarkan hasil penelitian dapat disimpulkan bahwa dari 54 dataset yang dilakukan pada tahap pengujian, akurasi mencapai 95% diperolah dalam menentukan kematangan dengan menggunakan LDA model warna *CIElab.*

Penelitian Selanjtunya dilakukan oleh(Susanto & Mulyono, 2022) dengan judul penelitian “Perbandingan Klasifikasi Jenis Apel Berkulit Merah Menggunakan Algoritma Linear Discriminant Analysis dan K-Nearest Neighbor”. Penelitian ini berfokus pada pengembangan sistem klasifikasi jenis apel berdasarkan citra kulit apel yang berwarna merah. Kendala dalam proses klasifikasi ini adalah kesamaan warna kulit dan bentuk apel yang hampir mirip antara jenis apel yang berbeda. Tujuan penelitian mengimplementasikan teknologi computer vision dengan menggunakan metode K Nearest Neighbor (KNN) dan Linear Discriminant Analysis (LDA) untuk mengklasifikasikan jenis apel berdasarkan citra kulit apel berwarna merah. Penelitian ini juga bertujuan untuk membandingkan hasil akurasi antara KNN dan LDA dengan dan tanpa ekstraksi fitur GLCM. Dataset yang digunakan terdiri dari 8 jenis apel merah. Berdasarkan hasil penelitian menunjukkan bahwa akurasi yang tinggi dalam klasifikasi tanpa menggunakan ekstraksi fitur GLCM, keduanya mencapai akurasi sekitar 99,25% dan 99%. Bahkan tanpa menggunakan fitur ekstraksi apapun, akurasi tetap tinggi sekitar 99,25% dan 99%. Dalam perbandingan antara KNN dan LDA, KNN memiliki sedikit keunggulan dalam akurasi. Dengan demikian, penelitian ini menunjukkan bahwa teknologi computer vision dengan menggunakan metode KNN dan LDA dapat efektif digunakan untuk mengklasifikasikan jenis apel berdasarkan citra kulit.

Penelitian oleh(Destriana et al., 2021) dengan judul “Implementasi Metode Linear Discriminant Analysis (LDA) Pada Klasifikasi Tingkat Kematangan Buah Nanas”. Permasalah pada penelitian pemilihan buah nanas dengan tingkat

kematangan yang tepat umumnya dilakukan secara manual, yang tidak efisien jika jumlah buah nanas yang besar perlu diseleksi. Penelitian mengembangkan sistem pengolahan citra untuk mengklasifikasikan tingkat kematangan buah nanas berdasarkan citra buah. Tujuan dari penelitian ini mengembangkan sistem pengolahan citra yang dapat mengklasifikasikan tingkat kematangan buah nanas berdasarkan ciri-ciri warna pada citra buah nanas. Berdasarkan hasil penelitian menunjukkan tingkat akurasi sebesar 83%, sistem digunakan untuk mengklasifikasikan tingkat kematangan buah nanas berdasarkan citra buah yang efisien.

Penelitian yang dilakukan oleh(Rizka Purmaya Sari, Ulla Delfana Rosiani, 2020) dengan judul peneltian “Implementasi Metode Linear Discriminant Analysis Untuk Deteksi Kematangan Pada Buah Stroberi”. Pada penelitian ini proses penyortiran buah stroberi secara manual oleh tenaga manusia dianggap kurang efektif karena dapat menyebabkan kesalahan dalam proses penyortiran. Berdasarkan hassil yang telah dilakukan penelitian menunjukkan bahwa metode Linear Discriminant Analysis (LDA) memiliki tingkat akurasi yang cukup baik dalam deteksi kematangan buah stroberi. Hasil akurasi metode ini adalah 85,71% untuk grade A, 84,28% untuk grade B, dan 97,14% untuk grade C. Dari keseluruhan dataset yang terdiri dari 210 dataset, sebanyak 177 dataset benar dan 43 dataset salah, dengan tingkat akurasi keseluruhan sebesar 84,28%.

Penelitian oleh(Sulaeman, 2022) dengan judul “Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke Support Vector Machine Algorithm Analysis In Stroke Disease Classification”. Tujuan dilakukan penelitian

untuk menguji algoritma SVM dalam klasifikasi data penyakit stroke. Berdasarkan hasil program machine learning klasifikasi data penyakit stroke dengan akurasi yang bervariasi pada jenis kernel yang digunakan. Hasil akurasi tertinggi diperoleh pada kernel polynomial akurasi 80%, sedangkan kernel linear sebesar 76%.

Selanjutnya penelitian oleh (Naufal et al., 2020) dengan judul penelitian "Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker dengan Data Microarray". Penelitian ini bertujuan untuk membangun sebuah sistem klasifikasi yang dapat digunakan untuk memprediksi kanker berdasarkan data microarray. Permasalahan dalam penelitian ini adalah dimensi yang sangat besar pada data microarray dan memilih metode klasifikasi yang paling akurat. Data yang digunakan empat jenis data yaitu data breast cancer, lung cancer, colon tumor, dan leukemia data tersebut diperoleh dari Kent Ridge Biomedical Data Repository. Berdasarkan hasil penelitian dapat disimpulkan bahwa proses reduksi dimensi sangat berpengaruh pada akurasi sistem klasifikasi dari keempat jenis data. Data leukemia mencapai akurasi tertinggi yaitu 98,54% dengan menggunakan metode klasifikasi KNN dengan k=5. Data breast memperoleh akurasi sebesar 66,52% dengan menggunakan klasifikasi KNN dengan k=9. Sedangkan data colon mencapai akurasi 85,60% dengan menggunakan klasifikasi SVM dengan kernel RBF, parameter C=1, dan gamma=1. Untuk data lung, akurasi tertinggi diperoleh tanpa melalui proses reduksi dimensi yaitu 100% dengan metode KNN dan k=5. Selain itu, kombinasi yang tepat antara penggunaan kernel dan parameter seperti C, d, dan gamma pada SVM juga dapat meningkatkan akurasi sistem klasifikasi yang telah dibuat.

Penelitian selanjutnya oleh (Deka Setia Negara, 2020) dengan judul penelitian "Perbandingan Algoritma Neural Network Dengan Linier Discriminant Analysis (Lda) Pada Klasifikasi Penyakit Diabetes Di Rsud Kudus". Salah satu permasalahan pada penelitian ini adalah penyakit Diabetes Melitus (DM) yang merupakan penyakit kronis dan mematikan, penyakit ini membutuhkan perawatan medis yang kontinu untuk mencegah komplikasi. Penelitian menggunakan algoritma Linear Discriminant Analysis (LDA) dan Neural Network untuk melakukan klasifikasi apakah seseorang terkena penyakit diabetes atau tidak. Data yang digunakan data pasien yang terkena penyakit diabetes. Data ini terdiri dari 520 sampel yang dibagi menjadi 416 data training dan 104 data testing untuk melatih dan menguji performa kedua algoritma. Berdasarkan hasil penelitian tingkat akurasi yang dicapai menggunakan Neural Network adalah 95,19%, sedangkan LDA mencapai akurasi 90,38%. Hal ini menunjukkan bahwa Neural Network lebih efektif dalam memprediksi penyakit diabetes berdasarkan data yang digunakan dalam penelitian tersebut.

Penelitian yang dilakukan oleh (Cahyanto et al., 2022) dengan judul "Pengujian Rule-Based Pada Dataset Log Server Menggunakan Support Vector Machine Berbasis Linear Discriminat Analysis Untuk Deteksi Malicious Activity". Tujuan penelitian ini adalah melakukan analisis log server menggunakan metode data mining untuk mengidentifikasi aktivitas yang mencurigakan atau berbahaya. Metode yang digunakan adalah Support Vector Machine (SVM) berbasis Linear Discriminant Analysis (LDA) dan pelabelan berbasis rule-based. Tujuan lainnya adalah menguji akurasi metode yang digunakan tersebut. Kesimpulan dari hasil

penelitian metode SVM berbasis LDA dengan pemodelan LDA+SVM-RBF mampu mencapai tingkat akurasi yang baik dalam melakukan pelabelan aktifitas pada web server. Dalam pengujian, metode tersebut memiliki tingkat akurasi rata-rata untuk training sebesar 98% dan untuk testing sebesar 84%.

Penelitian yang dilakukan oleh (Intan et al., 2022) dengan judul "Skin Disease Pattern Recognition Application Using Linear Discriminant Analysis Algorithm". Penelitian ini bertujuan untuk menjembatani antara pasien dan pemeriksaan penyakit kulit dengan menggunakan aplikasi berbasis kecerdasan buatan untuk mendiagnosa penyakit kulit. Metode yang digunakan adalah Linear Discriminant Analysis (LDA) berdasarkan objek citra untuk ekstraksi fitur dan klasifikasi penyakit kulit. LDA digunakan untuk membedakan ciri-ciri dalam kelas yang sama dan kelas yang berbeda, sedangkan Euclidean Distance digunakan untuk menghitung jarak antara fitur-fitur yang diekstraksi dengan fitur-fitur pada kelas yang telah ditentukan. Data yang digunakan dalam penelitian ini mencakup dokumen-dokumen yang berisi foto-foto yang diambil selama observasi, serta gambar yang disediakan di internet. Jumlah sampel penderita penyakit kulit yang digunakan adalah 50 orang, dan jumlah gambar yang digunakan adalah 66. Berdasarkan hasil penelitian, aplikasi yang dikembangkan mampu memberikan informasi awal bagi pasien mengenai jenis penyakit kulit yang dideritanya dengan tingkat akurasi sebesar 80%. Namun, peningkatan akurasi dapat dilakukan dengan menggunakan metode ekstraksi fitur lainnya dan penambahan dataset yang lebih proporsional antara data training dan data testing.

II.2 *Digital Image Processing*

(Vankdothu & Hameed, 2022)*Digital image processing* digunakan untuk memproses gambar dalam produksi gambar digital. Pemindai atau digitizer akan digunakan untuk mengonversi gambar ke format digital yang kemudian akan diproses. Pemrosesan citra digital adalah proses komputer digital memproses gambar dua dimensi. Ini mengacu pada pemrosesan digital dari setiap data dua dimensi dalam arti yang lebih besar. Sekelompok bilangan real diwakili oleh jumlah bit terbatas dalam gambar digital. Reproduktifitas, kemampuan beradaptasi, dan retensi data asli adalah manfaat utama dari sistem pemrosesan gambar digital.

Citra digital merujuk pada citra RGB adalah (gambar) yang memiliki nilai-nilai tetap untuk mengindikasikan intensitas warna gambar dari masing-masing pixel. Pixel sendiri adalah elemen pada citra digital yang direpresentasikan dengan angka numerik. Pada citra *grayscale*, intensitas pixel hanya direpresentasikan dengan satu nilai dan berkisar dari rentang nilai 0 – 255. Sedangkan pada citra berwarna, intensitas pixel direpresentasikan menjadi tiga nilai yang masing-masing mencerminkan nilai *Red, Green* dan *Blue* (RGB). *Digital image processing* melibatkan sejumlah teknik dan algoritma yang dapat dilakukan(Nugraha et al., 2022).

II.3 Citr*a*

Citra merupakan representasi objek dua dimensi dari dua dunia visual, menyangkut berbagai macam disiplin ilmu yang mencangkup seni human vision, astronomi, teknik, dan sebagainya. Citra kumpulan dari piksel-piksel atau titik-titik yang berwarna berbentuk dua dimensi. Citra komponen dari multimedia salah satu

bentuk informasi yang diperlukan manusia selain teks, suara dan video. Informasi yang terkandung dalam sebuah citra dapat diinterpretasikan berbeda-beda oleh manusia satu dengan yang lain. Pengolahan citra adalah suatu jenis teknologi untuk menyelesaikan masalah mengenai pemrosesan gambar, sedangkan computer vision mempunyai tugas untuk membuat suatu keputusan tentang objek fisik nyata yang didapat dari perangkat atau sensor(Pratama, 2022).

II.4 Citra *Grayscale*

Citra *grayscale* yaitu citra yang nilai pikselnya merepresentasikan derajat keabuan atas intensitas warna putih. Setiap nilai piksel di dalam citra *grayscale* sesuai denga kecerahannya. Nilai piksel citra *grayscale* akan direpresentasikan oleh byte atau *word* dengan nila 8-bit, intensitas kecerahan yang berfariasi dimulai dari 0 sampai 255, “0” direpresentasikan sebagai hitam dan “255” direpresentasikan sebagai putih(Faisal et al., 2019).

Grayscale = ((R * 0.2989) + (G * 0.5870) + (B * 0.1140))

II.5 Citra Digital

Secara umum citra digital merupakan hasil pengolahan dari komputer berupa gambar atau image yang memiliki nilai atau piksel yang diperoleh secara matematis sehingga menghasilkan gambar dua dimensi atau gambar f(x,y), di mana bayangan f(x,y) adalah cahaya intensitas dengan nilai x dan nilai y yaitu titik tingkat kecemerlangan. Citra digital menghasilkan a matriks yang indeks kolom dan barisnya pada setiap titik mewakili tingkat abu-abu (Marchetti et al., 2022).

Citra atau disebut juga dengan gambar adalah fungsi *continue* yang memiliki nilai intensitas cahaya dua dimensi. Dan digital adalah pengolahan citra menggunakan komputer yang dilakukan secara digital. Citra harus direpresentasikan secara numerik dengan nilai diskrit agar dapat di oleh dengan mudah dan proses tersebut disebut denga digitalisasi citra. Digitalisasi citra adalah representasi dari fungsi *continue* menjadi nilai-nilai diskrit. Terdapat sebuah matriks dalam sebuah ciitra digital yang mempunyai dua dimensi yaitu f(x,y) yang terdiri dari kolom dan baris. Piksel merupakan perpotongan antara baris dan kolom(Laksono & Rusmamto, 2022).

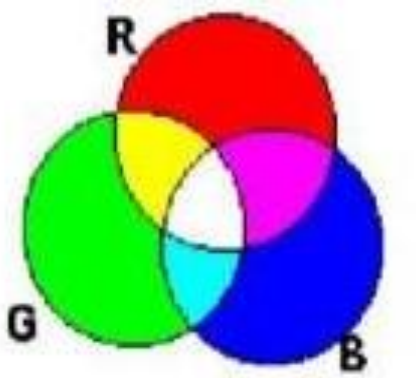
II.6 Binary Image

Binary image atau disebut dengan citra biner merupakan citra yang setiap pikselnya memiliki 2 warna yaitu warna hitam atau warna putih, dan hanya terdapat dua warna pada setiap piksel dan diperlukan 1 (satu) bit atau 8 (delapan bit) per piksel, sehingga penyimpanan dapat dilakukan secara efisien. Citra biner berguna untuk penulisan teks, fingerprint dan arsitektur gambar(Laksono & Rusmamto, 2022).

II.7 *Red, Green* dan *Blue* (RGB)

Red, Green, Blue (RGB) merupakan ruang warna yang sangat dikenal dan banyak digunakan dalam menghasilkan citra. Model warna RGB adalah aditif model warna yang merah, hijau dan biru ditambahkan bersama dalam berbagai cara untuk mereproduksi array yang lebih luas dari warna. Tujuan utama RGB adalah

untuk sensing, representasi, dan menampilkan gambar dalam sistem elektronik. Perangkat elektronik seperti kamera, perangkat penampil televisi memang menggunakan ruang warna ini karena bisa menghasilkan warna-warna dasar dan kombinasi yang bagus. Hal ini dikarenakan mata memang mampu memberikan persepsi warna yang kuat terhadap ruang warna ini. Namun menjadi kendala ketika ingin menginterpretasikan warna yang nyata dan tidak termasuk ke dalam ruang warna RGB. Jika melihat warna *cyan* atau magenta atau warna lainnya yang bukan warna dasar RGB maka akan kesulitan dalam menjelaskan warna-warna apa saja yang menyusun warna tersebut. Jika 10% hijau, 30% biru dan 60% merah. Tentu hal itu sangat tidak praktis, dan belum tentu akurat. Lebih lagi, mata lebih mudah untuk mempersepsikan warna itu dalam istilah kecerahannya (Ninosaria & Mardiana, n.d.). Untuk itu dibutuhkan ruang warna lain yang bisa membantu dalam menjelaskan ciri warna yaitu HIS.



Gambar II.1 Komposisi Warna RBG

II.8 Pre-processing

Pengolahan citra digital merupakan bentuk pemrosesan atau pengolahan sinyal dengan masukan beberapa citra (image) dan ditrasformasikan menjadi citra lain sebagai keluarannya dengan menggunakan Teknik-teknik tertentu. Input dari pengolahan citra adalah citra, sedangkan outputnya adalah citra hasil pengolahan(M & Azizah, 2022).

II.9 Ekstraksi Citra

Ekstraksi citra digital adalah proses pengindeksan suatu database berupa citra (image) dengan isinya. Salah satu proses ekstraksi fitur adalah menganalisa berdasarkan isi visual seperti warna, bentuk dan tekstur. Setelah proses ekstraksi fitur, dilakukan proses klasifikasi untuk menentukan tingkat keakuratan dari proses identifikasi yang telah dilakukan. Pengelolaan citra digital memiliki beberapa tujuan, salah satunya untuk mengambil ciri dari citra sehingga dapat dikenali yang biasa disebut ekstraksi dengan ciri (Syahri, 2023).

II.9.1 Ekstraksi Fitur Warna dan Bentuk

(Ge et al., 2022)Fitur warna dan bentuk digunakan untuk klasifikasi citra buah sawit dan pengenalan buah sawit. Fitur warna, terdiri dari nilai RGB, nilai HSV dan fitur bentuk terdiri dari Cur-vatures, Fast Point Feature Histogram (FPFH), Spin Image. Fitur-fitur ini dipilih akhirnya, melalui banyak pra-pelatihan dan pra-tes. Untuk set pelatihan, buah sawit diklasifikasikan secara manual, dan label yang sesuai dibuat pada saat yang sama.

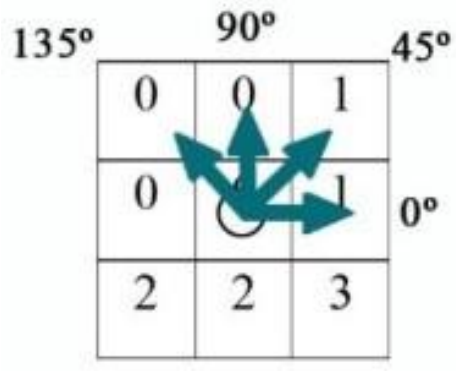
II.10 Segmentasi Citra

Image segmentation atau segmentasi citra adalah sebuah proses untuk memisahkan sebuah objek dari background, sehingga objek tersebut dapat diproses untuk keperluan lain. Dengan proses segmentasi tersebut masing-masing objek pada gambar dapat diambil secara individu sehingga dapat digunakan sebagai input bagi proses segmentasi untuk memisahkan objek yang akan direkontruksi terhadap background yang ada (Pangaribuan & Sitohang, 2023).

Pada penelitian(Yunianto et al., 2023) menggunakan metode *otsu thresholding* segmentasi gambar digunakan untuk melibatkan pemisahan objek dari item lain atau latar belakang dari gambar. Metode ambang batas memiliki nilai ambang batas yang digunakan untuk mengonversi gambar skala abu-abu menjadi bentuk biner.

II.11 *Gray Level Co-occurrence Matrix* (GLCM)

(Songket et al., 2020)GLCM pertama kali diperkenalkan dengan nama *Gray-Tone Spatial-Dependence Matrix*. GLCM adalah matriks yang menggambarkan frekuensi munculnya pasangan dua piksel dengan intensitas tertentu dalam jarak d dan orientasi arah dengan sudut θ tertentu dalam citra. Jarak dinyatakan dalam piksel, biasanya 1, 2, 3 dan seterusnya. Orientasi sudut dinyatakan dalam derajat, yaitu 0°, 45°, 90° dan 135°. Adapun arah sudut dalam matriks GLCM dapat dilihat pada Gambar II.3



Gambar II.2 Arah dalam GLCM

(Annas Prasetio et al., 2022)GLCM adalah teknik untuk memperoleh distribusi derajat keabu-abuan (histogram) dengan nilai statistik orde 2 dengan menghitung pengukuran derajat kontras, granularitas, dan probabilitas kedekatan antara dua piksel pada kekasaran tertentu dari suatu area hubungan efisiensi antara piksel dalam gambar.

Gray Level Co-occurrence Matrix (GLCM) adalah salah satu metode ekstraksi fitur pada tekstur dan termasuk dalam statistik orde kedua sebagai teknik untuk memperoleh nilai dengan menghitung probabilitas hubungan tetangga atau kedekatan dua potong (Pulung Nurtantio Andono1, 2022). piksel pada jarak yang telah ditentukan (d) dan sudut (θ). Arah Sudut dalam GLCM Hasil yang dapat diperoleh dari perhitungan metode GLCM antara lain:

Contrast: Adalah perhitungan yang berkaitan dengan jumlah keragaman intensitas gambar skala abu-abu.

$$\text{Con} = \sum_x \sum_y (x - y)^2 \, p(x,y) \qquad (1)$$

Correlation: Memberikan petunjuk tentang keberadaan struktur linier pada gambar dengan menunjukkan ukuran ketergantungan linier derajat abu-abu pada

gambar.

$$Cor = \sum_x \sum_y (\_\_\_x - \mu x)(\sigma x y - \sigma y\ \mu y)\ p(x,y) \qquad (2)$$

Energy: Tentukan intensitas keabuan dengan ukuran konstransi pasangan tertentu pada matriks.

$$Eng = \sum_x \sum_y p(x, y)^2 \qquad (3)$$

Homogeneity: Jumlah level abu-abu piksel gambar akan lebih tinggi jika memiliki nilai yang sama. Ini juga berlaku untuk nilai terbalik GLCM.

$$Hom = \sum x \sum y\ 1\_\_\_\_\_ p+(\ |xx,y-)y| \qquad (4)$$

Penjelasan x= Value line dalam matriks kookurensi, y=Value kolom dalam matriks kookurensi p (x,y) = Nilai garis (x) dan kolom (y) pada matriks kookurensi:

μx = average of line value (x) = $\sum x \sum y\ xp(x,y)$

μy = average of column value (y) = $\sum y \sum y\ yp(x,y)$

σx = Variance matrik

(X) = $\sqrt{\sum_x \sum_y (x - \mu x)^2\ p(x,y)}$

σy = Variance matrik

(Y) = $\sqrt{\sum_x \sum_y (y - \mu y)^2\ p(x,y)}$

Hasil data tersebut akan digunakan sebagai data dari klasifikasi algoritma untuk menemukan kelas suatu objek pada gambar.

II.12 Klasifikasi

Klasifikasi adalah suatu proses pengelompokan data yang didasarkan pada ciri-ciri tertentu kedalam kelas yang sudah ditentukan terlebih dahulu. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Teknik klasifikasi bekerja dengan mengelompokkan data berdasarkan data training dan nilai atribut klasifikasi (Affandi et al., 2022).

Dalam klasifikasi data set yang digunakan harus memiliki label atau atribut tujuan. Meramal objek kelas pada setiap persoalan dalam data adalah tujuan dari klasifikasi. Dimulai dengan satu set data dimana kelas dikenal adalah sebuah tugas klasifikasi (Maisa Hana et al., 2023).

(Salim & Suryadibrata, 2019)Berdasarkan data training yang digunakan, klasifikasi terbagi atas 2 jenis:

1. *Supervised Classification* Merupakan proses menggunakan data yang telah diketahui untuk mengklasifikasikan data asing. Contoh algoritma: Logistic Regression, Naïve Bayes, SVM (Support Vector Machine), ANN (Artificial Neural Networks), dan Random Forests.
2. *Unsupervised Classification* Merupakan proses clustering, yaitu pembagian data asing dalam jumlah tertentu menjadi beberapa kelompok berdasarkan fitur yang serupa. Contoh algoritma: K-Means Clustering, PCA (Principal Component Analysis), dan Autoencoder.

II.13 Tanaman Kelapa Sawit

Nama latin kelapa sawit *elates guineensis* berasal dari Bahasa Yunani kuno *elola* yang berarti *zaltun*. Nama diberikan karna mengandung buah minyak. Kelapa sawit memiliki organ vegetative berupa daun, batang, akar, serta organ, reproduktif berupa bunga dan buah(Pt & Prima, 2022).

II.13.1 Penyakit Tanaman Sawit

Penyakit tanaman sawit adalah penyakit atau gangguan dari hama yang dimiliki setiap tanaman sawit sehingga tanaman sakit dan tumbuh tidak dengan optimal. Tetapi petani sering mengabaikan hal ini karna ketidaktahuaanya dan menganggap gejala tersebut sudah biasa terjadi pada masa tanam, sampai suatu saat timbul gejala yang sangat parah dan meluas, sehingga sudah terlambat untuk dikendalikan(Pt & Prima, 2022).

II.14 Linear *Discriminant Analysis*

Linear Discriminant Analysis (LDA) pertama kali diterapkan oleh Etemad dan Chellapa pada proses pengenalan wajah. Teori stastiktik merupakan metode yang digunakan oleh LDA untuk machine learning pengolahan data dan pengolahan citra. Pendekatan LDA dengan mengenali pola tertentu dalam menemukan kombinasi fitur berdasarkan pembelajaran. Analisa matriks penyebaran (*scatter matrix analysis*) merupakan cara kerja LDA dengan tujuan menemukan proyeksi optimal untuk memaksimalkan penyebaran antara kelas dan meminimalkan penyebaran dalam kelas data buah sawit(Laksono & Rusmamto, 2022).

Algoritma *Linear Discriminant Analysis* (LDA) merupakan salah satu *supervised learning* dalam pengolahan citra yang lebih dikenal sebagai *Fisher's Linear Discriminant,* metode ini ditemukan oleh Ronal A. Fisher pada tahun 1936. LDA adalah metode ekstraksi fitur dengan perpaduan dari perhitungan operasi matematika dan statistika yang memberlakukan property statistic terpisah untuk tiap obyek. Tujuan metode LDA adalah mencari proyeksi linear (yang biasa disebut dengan '*fisherimage*') untuk memaksimumkan matriks kovarian antar kelas (*between-classs covariance matrix*), agar anggota didalam kelas lebih terkumpul penyebarannya dan pada akhirnya dapat meningkatkan keberhasilan pengenalan(Intan et al., 2022) [7].

LDA memilih fitur yang memaksimalkan rasio dari antar kelas dan menyebar kedalam kelas, didefinisikan sebagai:

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^r$$

Sw = Matriks kovarian dalam kelas

SB = Matriks kovarian antar kelas

C = jumlah kelas

Ni = jumlah image pada kelas ke-i

u = rata-rata dari total keseluruhan image

ui = rata-rata image pada kelas-i

Dimana u rata-rata total dari keseluruhan image dan mean adalah rata-rata image pada kelas ke-i. SB yang dimaksud adalah *between class scatter matrix* dan *within class scatter matrix* Sw didefinisikan sebagai berikut:

$$Sw = \sum_{\varepsilon=1}^{c} P_l S_1$$

dimana

$$S_1 = E[(\mu_1 - \mu)(\mu_1 - \mu)^T | X \in C_1]$$

jika Sw non-tunggal, proyeksi optimal dipilih dari matriks dengan kolom ortonormal yang memaksimalkan rasio dari determinan *between class scatter matrikx* dengan determinan dari *within class scatter matrix*. Dimana Wopt adalah:

$$J(Wopt) = \arg\max_w \frac{|w^T SbW|}{|w^T SwW|}$$

$$SbW_1 = A_l SwW_1$$

Kebanyakan C-1 *nonzero* menyimpulkan nilai *eigen*, dimana C adalah jumlah kelas. *Linear Discriminant Analysis* (LDA) secara luas digunakan untuk menemukan kombinasi *linear* fitur sambil menjaga kelas keterpisahan. Telah diketahui bahwa tujuan dari LDA adalah untuk memaksimalkan *between class scatter matrix* sekaligus meminimalkan tersebarnya di dalam kelas. LDA sebagai Teknik reduksi data yang diawasi, bertujuan untuk mengekstraksi fitur yang memaksimalkan keterpisahan di antara kelas-kelas yang berbeda dalam data.

II.15 *Support Vector Machine*

(Marchetti et al., 2022)*Support vector machine* adalah algoritma pembelajaran mesin yang awalnya dikembangkan oleh Vapnik dan rekan kerjanya Vladimir, dan berturut-turut diimplementasikan/diperluas untuk menciptakan apa yang disebut metode berbasis kernel. Dalam klasifikasi, SVM menghasilkan fungsi dengan belajar dari dataset pelatihan, untuk menemukan hyperplane terbaik yang memaksimalkan jarak di antara berbagai kelas data.

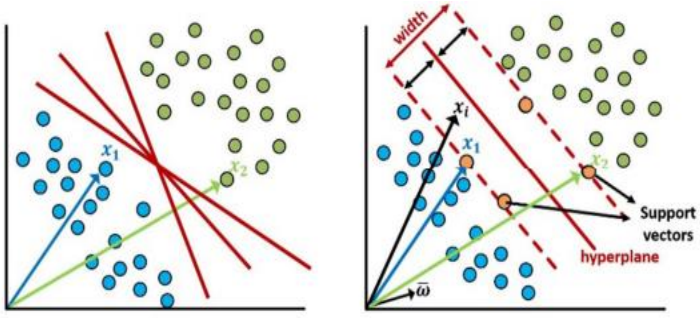
(Vankdothu & Hameed, 2022)*Support Vector Machine* adalah teknik

pembelajaran yang diawasi dari bidang materi pembelajaran mesin hingga klasifikasi dan regresi. Target dari pendekatan SVM adalah untuk menemukan hyperplane yang meningkatkan tepi geometris dan membatasi kesalahan klasifikasi ketika diberi masalah *line-arly* dibedakan dua kelas. SVM dipandang sebagai sistem yang berharga untuk klasifikasi *nonlinier*. Dengan tujuan akhir tertentu untuk memainkan proses non-langsung, fungsi kernel dimulai dalam klasifikasi SVM. Dua tahap kunci yang tidak penting dari prosedur SVM adalah tahap pelatihan dan tahap pengujian. Mengingat pengaturan ilustrasi persiapan masing-masing dipisahkan sebagai memiliki tempat dengan salah satu dari dua kelas, algoritma pelatihan SVM menghasilkan model yang menurunkan kasus baru ke dalam satu klasifikasi atau yang lain.

Support Vector Machine merupakan sekumpulan metode *supervised learning* yang memuat *hyperlane* atau sekumpulan *hyperlane* pada proses klasifikasi, regresi, dan *outlier detection*. Kelebihan pada SVM ini adalah efektifitas pada *high dimensional space*, efektif dalam kasus dengan jumlah dimensi yang lebih banyak dari pada jumlah sampelnya, menggunakan subset titik pelatihan sehingga memori lebih efisien(Pratama, 2022). Suatu *support vector machine* melakukan klasifikasi dengan membangun sebuah *hyperplane* N-dimensi yang optimal memisahkan data menjadi dua kategori. Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space*. *Pattern* yang merupakan anggota dari dua buah kelas: +1 (kuning) dan 1 (merah), dan berbagai alternative baris pemisah (*discrimination boundaris*). Margin adalah jarak antara *hyperplane* tersebut dengan *pattern* terdekat dari

masing-masing kelas. *Pattern* yang paling dekat ini disebut sebagai *support vector*(Data et al., 2022).

Algoritma SVM mengklasifikasikan data dengan mencari *hyperplane* terbaik yang memisahkan semua data dari satu kelas atau beberapa kelas. SVM mendekati masalah melalui konsep jarak terkecil antara batas keputusan dan salah satu sampel data(Ricky Yohannes1, 2022). Dapat dilihat pada Gambar II.5.



Gambar II.3 Hyperlane Support Vector Machine

Support Vector Machine adalah pengklasifikasi yang memisahkan sampel ke dalam kelompok-kelompok yang berbeda berdasarkan konsep *hyperplane*. Pembelajaran SVM akan menghasilkan *hyperplane* atau pembatas pemisahan antara dua kelas berdasarkan sifat data. Dalam proses pelatihan, SVM berusaha mencari gradien *hyperplane* yang optimal. Secara teoritis, pengklasifikasi SVM tradisional yang menggunakan pendekatan satu lawan satu hanya dapat melakukan pemisahan data kedalam dua kelas yang berbeda sekaligus, yang tidak kompatibel dengan klasifikasi multi-kelas (Ong, J.H., Ong, P., & Woon, 2022). Fungsi karnel digunakan untuk memetakan data yang tidak dapat dipisahkan secara *linear* (misalnya citra penginderaan jauh) ke dalam ruang dimensi yang lebih tinggi untuk mendefinisikan *hyperplane* yang optimal. Fungsi karnel yang biasa digunakan

dalam SVM dapat dikelompokkan secara umum ke dalam kelompok, yaitu karnel *linear*, *polinomial*, *radial basis*, dan s*igmoid*(Kavzoglu & Colkesen, 2012).

Berikut ini merupakan kekuatan dari Support Vector Machine (SVM)(Hermanto, Ali Mustopa, 2020) antara lain:

1. Mempunyai kemampuan generalisasi yang tinggi.
2. Mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relatif sedikit hanya dengan pengaturan parameter yang sederhana. SVM memiliki konsep dan formulasi yang jelas dengan sedikit parameter yang harus diatur.
3. Relatif mudah diimplementasikan karena penentuan SVM dapat dirumuskan dalam masalah QP (Quadratic Programming).

Sementara itu, kelemahan yang terdapat dalam Support Vector Machine (SVM) sebagai berikut:

1. Sulit diaplikasikan untuk himpunan data dengan jumlah sampel dan dimensi yang sangat besar.
2. Umumnya hanya diformulasikan untuk meyelesaikan masalah klasifikasi dua kelas. Walupun dapat dikembangkan untuk menyelesaikan masalah klasifikasi multi kelas, namun masing-masing strategi multi kelas SVM juga memiliki kelemahan.

II.16 *Confussion Matrix*

Confussion matrix yaitu merupakan pengujian untuk mengetahui kinerja algoritma pembelajaran mesin (Handayani et al., 2021). *Confussion matrix* adalah suatu metode yang memberikan informasi hasil klasifikasi yang dilakukan oleh sistem untuk menganalisa seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda(Apriyani, 2020). Data yang diuji tingkat akurasinya menggunakan *confussion matrix*. Metode *confussion matrix* digunakan sebagai pengukur proses suatu metode klasifikasi. *Confussion matrix* mengandung empat istilah yaitu *True Positive* (TP), *True Negative* (TN), *False Posetive* (FP), *False Negative* (FN)(Amelia et al., 2022). Tabel II.1 merupakan dasar untuk melakukan perhitungan *Confussion Matrix* sebagai berikut:

Tabel II.1 Metode *Confussion Matrix*

		Observed	
		True	***False***
Predicted Class	***True***	*True Positive* (TP)	*False Positive* (FP)
Predicted Class	***False***	*False Negative* (FN)	*True Negative* (TN)

TP adalah akumulasi data positif yang terklasifikasi dengan tepat oleh sistem. Kemudian TN adalah akumulasi data negatif yang terklasifikasi dengan tepat oleh sistem. FN adalah akumulasi data negatif yang terklasifikasi salah oleh sistem. FP adalah akumulasi data positif yang terklasifiksi salah oleh sistem.

Confussion matrix terdapat dari beberapa perhitungan jumlah klasifikas yaitu:

1. Presisi

Presisi merupakan perhitungan untuk *mengetahui* jumlah data yang benar positif dari semua hasil prediksi benar positif. Presisi dapat dilakukan menggunakan persamaan.

$$Precision = \frac{TP}{TP + FP}$$

2. Recall

Recall merupakan perhitungan untuk mengetahui jumlah data yang prediksi benar positif dari hasil seluruh benar positif. Recall menggunakan persamaan

$$Recall = \frac{TP}{TP + FN}$$

3. F1 *Score*

F1 *Score* merupakan perhitungan untuk mengetahui rata-rata dari perbandingan presisi maupun recall. F1 Score dapat dilakukan perhitungan menggunakan persamaan

$$F1\ Score = \frac{Precision\ x\ Recall}{Precision +\ Recall}$$

4. Akurasi

Akurasi merupakan perhitungan untuk mengetahui keakuratan model dalam klasifikasi yang benar. Akurasi dapat dilakukan dengan persamaan.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

II.17 Penelitian Sebelumnya

Adapun beberapa penelitian atau jurnal yang menjadi acuan dalam penelitian ini dapat dilihat pada Tabel II.1

Tabel II.2 Penelitian sebelumnya

No.	Penelitian/ TahunTerbit	Judul Penelitian	Persamaan	Perbedaan
1.	(Marchetti et al., 2022)	Klasifikasi Konsentrat Bijih Uranium Menerapkan Support Vector Machine untuk fitur spektrofotometri dan tekstur	Sama-sama menggunakan support vector machine pada proses klasifikasi	Tidak menggunakan LDA
2.	(Yunianto et al., 2023)	Analisis Kinerja Komparatif Deteksi Kanker Paru menggunakan Naïve Bayes, Support Vector Machine, K-Nearest Neighbor dan Decision Tree	Sama-sama menggunakan analisis kinerja support vector machine	Tidak menggunakan Naïve Bayes, K-Nearest Neighbor dan Decision Tree
3.	(Anzid et al., 2019)	Klasifikasi Gambar Multimodal menggunakan Dense SURF, Spectral Information dan Support Vector Machine	Sama-sama menggunakan Klasifikasi Gambar dengan SVM	Tidak menggunakan Dense SURF dan Spectral Information
4.	(Jana et al., 2023)	Analisis fitur klasifikasi jenis anggur	Sama-sama menggunakan	Tidak menggunakan

		berdasarkan kualitas melalui jaringan saraf dan support vector machine	metode klasifikasi support vetor machine	LDA
5.	(S. Yang et al., 2018)	Deteksi Dini Penyakit Menggunakan Rekam Kesehatan Elektronik dan Analisis Discriminant Linear Fisher's Wishart	Sama-sama mengugunakan metode LDA	Tidak menggunakan metode SVM
6.	(Djoufack Nkengfack et al., 2021)	Perbandingan metode ekstraksi fitur PCA, KPCA, LDA dan GDA berbasis polinomial untuk deteksi sinyal EEG epilepsi dan mata menggunakan mesin kernel	Sama-sama mengugunakan metode LDA	Tidak menggunakan metode PCA,KPCA dan GDA
7.	(Ge et al., 2022)	Klasifikasi organ pohon apel tiga dimensi dan algoritma estimasi hasil berdasarkan multi- Fitur Fusion dan Support Vector Machine	Sama-sama menggunakan support vector machine	Tidak menggunakan Multi-Fitur Fusion pada klasifikasi pohon apel
8.	(Ribeiro et al., 2021)	FT-NIR dan analisis diskriminan linier untuk mengklasifikasikan benih buncis yang diproduksi dengan bahan kimia bantuan	Sama-sama analisis linear discriminant analyis	Tidak menggunakan metode SVM

		panen		
9	(Wei et al., 2022)	Analisis Diskriminan Linier Berdasarkan Metode Penentuan Posisi Untuk Pengukuran Jarak Isometrik Pemotongan Laser Benih Jagung	Sama-sama menggunakan Linear Discriminant Analyis pada klasifikasi objek	Tidak menggunakan metode K nearest neighbour clustering

II.18 Kontribusi Penelitian

Adapun kontribusi dalam penelitian ini adalah membandingkan *performance* antara algoritma LDA dan SVM untuk klasifikasi penyakit pada buah sawit khususnya terkait dengan variasi pada ukuran dataset dan juga persentase data *training* dan data *testing*